Biographies

**Katy Börner** is an Associate Professor of Information Science in the School of Library and Information Science, Adjunct Associate Professor in the School of Informatics, Core Faculty of Cognitive Science, Research Affiliate of the Biocomplexity Institute, Member of the Advanced Visualization Laboratory and directs the Information Visualization lab at Indiana University. Her research focuses on the development of data analysis and visualization techniques that improve information access, understanding, and management. She is particularly interested in the study of the structure and evolution of scientific disciplines; the analysis and visualization of online activity, e.g., user actions in 3D virtual worlds; and the development of cyberinfrastructures for scientific collaboration and computation, e.g., the Information Visualization Cyberinfrastructure (http://iv.slis.indiana.edu/), the Network Workbench (http://nwb.slis.indiana.edu), and the Scholarly Database (https://iv.slis.indiana.edu/db). She co-edited a book on 'Visual Interfaces to Digital Libraries' published by Springer in 2002, a special issue of PNAS on 'Mapping Knowledge Domains' published in April 2004, a special issue on 'Collaborative Information Visualization Environments' in PRESENCE: Teleoperators and Virtual Environments, MIT Press that appeared in Feb. 2005, and a special issue on 'Information Visualization Interfaces for Retrieval and Analysis' in the Journal of Digital Libraries that appeared in March 2005. A special issue of Environment and Planning B on 'Mapping Humanity's Knowledge and Expertise in the Digital Domain' is in preparation. Börner is the recipient of many fellowships and awards, including Outstanding Junior Faculty Award, Pervasive Technology Laboratories Fellowship, SBC Fellow, NSF CAREER Award, and Trustees Teaching Award. She is currently PI or Co-PI on 12 grants that are funded by NSF, the James S. McDonnell Foundation, 21st Century Fund, and SUN Microsystems.

Homepage: [http://ella.slis.indiana.edu/~katy/](http://ella.slis.indiana.edu/~katy/)
Relevant Publications: (visit [http://ella.slis.indiana.edu/~katy/cv](http://ella.slis.indiana.edu/~katy/cv) for full list)


Holloway, Todd, Božicevic, Miran and Börner, Katy. Analyzing and Visualizing the Semantic Coverage of Wikipedia and Its Authors. Accepted for *Complexity*, Special Issue on *Understanding Complex Systems*. Also available as [cs.IR/0512085](http://arxiv.org/abs/cs.IR/0512085).


**General Questions**

What is your main interest in attending the workshop?

I am interested to study the structure and evolution of our collective knowledge by scientific means. I believe we need to develop better information access and management tools or we are doomed to reinvent the wheel forever. Global science (weather) forecasts might help people understand and use what we know. The quality of any scientific analysis or map strongly depends on the quality of data available. I am interested to use high coverage scholarly data with unique document, author, institution, and geolocation information.

What is your main interest in ‘data sharing’ and/or ‘data integration’?

This workshop brings together some of the major scholarly data/service providers. It is my hope that we can collectively arrive at standards and procedures for (sharing) unique author, institution, and geolocation identifiers. We definitely need those for interlinking publications, patents, grants, etc. in our scholarly database (https://iv.slis.indiana.edu/db).

**Scholarly Datasets and Data Integration Challenges**

Please provide information on the datasets you serve/use in your research.

- **Dataset/Project Description**
  
  Scholarly Database (https://iv.slis.indiana.edu/db).

- **Preservation – what processes or procedures are in place to keep data useable over time?**
  
  None.

- **Provenance – what processes or procedures are in place to insure that the data contributed to the project is authentic and authoritative?**
  
  None.

- **Privacy Policy – what policies are in place to insure the privacy of contributors and/or end researchers?**
  
  Some datasets are proprietary. Most are freely available.
Data formats / types – what type of data is stored and in what format?

Database format.

Database Technology – what database management system is used (Oracle, MySQL, etc)?

PostgreSQL.

Storage Technology – what hardware is used to store the data?

Sun servers with 10TB hard disk space.

Backup Strategy – with what frequency is the data backed up? Are copies kept off site? Do you have a disaster recovery plan?

Indiana University mass storage and tape backups.

Partners – what other organizations participate in this project?

Kevin Boyack, Sandia National Laboratories.

Funding / Ownership – which organization owns this project, and how is this project funded?

NSF funding.

What data integration issues are you facing/addressing in your research? Please address policy and technical level.

Unique document, author, institution, and geolocation identifiers are needed.

What data integration solutions do you use/have you developed?

Simple lookup tables for mapping addresses to geolocations. Fuzzy matches for author names.

Please provide references/links/pointers to relevant work, papers, and efforts.

Please send the completed document by **August 22, 2006** to Katy Borner <katy@indiana.edu> and Stacy T. Kowalczyk <skowalcz@indiana.edu> using the subject header ‘ScholarlyDB&I’.